

WOODHOUSE EXHIBIT 14

EXHIBIT M

Message

From: Brian Gamido [REDACTED]@meta.com]
Sent: 3/14/2024 4:22:59 AM
To: Mike Clark [REDACTED]@meta.com]; Juan Cocuy [REDACTED]@meta.com]; Didem Foss [REDACTED]@meta.com]; Brian Gamido [REDACTED]@meta.com]; Guan Pang [REDACTED]@meta.com]; James Kohli [REDACTED]@meta.com]; Chaya Nayak [REDACTED]@meta.com]; Shelley Venus [REDACTED]@meta.com]; Sean Bell [REDACTED]@meta.com]; Kenneth Heafield [REDACTED]@meta.com]
Subject: Message summary [{"otherUserFbId":null,"threadFbId":7247680588621055}]
Attachments: 363519422_219978361029841_5840576661070952203_n.gif;
431597880_3770001743253513_7986028539968176578_n.png;
418596861_367281262805262_625167500481097423_n.png

James Kohli (3/13/2024 13:34:17 PDT):
>hey peeps. we've been asked to pull together a timeline on data sourcing across modalities (basically to do what we've done here for multiple data types)
>
>I'm thinking we divide and conquer on this, where PM can own 1P, partnerships can own potential licenses, and eng can own crawl/open targets.
>
>I've created a matrix here to get date estimates on these. Asks: can you a. sub in the correct owner if you're not the best one where your name is b. suggest some names for audio owners?

James Kohli (3/13/2024 13:35:10 PDT):
><https://docs.google.com/document/>[REDACTED]
#shareadoc edit

James Kohli (3/13/2024 13:35:33 PDT):
>whoops and c. provide an estimated timeline where your name is, ofc

James Kohli (3/13/2024 13:35:46 PDT):
>(also feel free to challenge this structure or let me know if you need more details)

Kenneth Heafield (3/13/2024 14:04:32 PDT):
>What is a list of targets for 1P text anyway?

Guan Pang (3/13/2024 14:05:17 PDT):
>same question for video/audio

James Kohli (3/13/2024 14:05:30 PDT):
>What's all the possible teams we could go talk to, essentially

James Kohli (3/13/2024 14:05:49 PDT):
>Maybe not very long? But there are quite a few surfaces across the company

Kenneth Heafield (3/13/2024 14:06:42 PDT):
>My understanding is we already have initial experiments done with the 1P data

James Kohli (3/13/2024 14:06:58 PDT):
>In fact, I think we may have created this already, give me a moment

Guan Pang (3/13/2024 14:07:02 PDT):
>It's good to list out all the domains, but we don't necessarily have to talk to them unless there's legal/privacy issue involved

Guan Pang (3/13/2024 14:08:07 PDT):
>1P text: FB, IG, ASR on all video, BizMsg. The last one needs product team support due to privacy.

James Kohli (3/13/2024 14:08:25 PDT):
><https://docs.google.com/spreadsheets/>[REDACTED] #shareadoc

Kenneth Heafield (3/13/2024 14:08:30 PDT):
>Also, each of these cells is a bundle of 20-300 different sources. For licensing probably on their own timelines. For free data sets, my primary question back is: am I the only one doing it or do I have people helping me download?

James Kohli (3/13/2024 14:08:32 PDT):
>The 1P tab

James Kohli (3/13/2024 14:09:31 PDT):
>Let's discuss!

James Kohli (3/13/2024 14:10:18 PDT):

>Maybe by mid next week we can estimate acquisition timelines with different resourcing levels?

Kenneth Heafield (3/13/2024 14:11:02 PDT):

>Like Shelley and Didem's shoes here: [REDACTED] is a "friendly" on a much shorter timeline than a random partner they've never worked with.

Guan Pang (3/13/2024 14:11:26 PDT):

>Amazing, for image/video we're already using the major FB/IG ones, but it's good to assess possibility/value of others. Like I have always been wondering if Ads media is something we can use.

Kenneth Heafield (3/13/2024 14:12:20 PDT):

>Hahaha lol I'm envisaging a model trained solely on FB ads and thinking it might actually be worth something

Guan Pang (3/13/2024 14:13:26 PDT):

>I'm pretty sure Sean has some model like that before he moved to GenAI😊

James Kohli (3/13/2024 14:21:16 PDT):

>To be clear though, we're not providing a number on how long it would take to get the data in this matrix, it's an estimate for how long it would take to put a timeline on acquiring each source (which they've done for studios already)

Shelley Venus (3/13/2024 16:05:54 PDT):

>@James Kohli are you still aiming to try and fill in this table in the larger data review doc
<https://docs.google.com/document/> [REDACTED]

Chaya Nayak (3/13/2024 16:06:56 PDT):

>yes please ASAP if you all can :)

Shelley Venus (3/13/2024 16:23:14 PDT):

>thanks @Chaya Nayak - just to clarify, when you say feasibility completed - does this mean we have completed an internal decision/escalation?

>
>For ex. I can tell you w/in a few days a solid cost assessment of news content, but it prob requires an internal escalation to MZ and Clegg to sign off which could take a bit longer

Chaya Nayak (3/13/2024 16:24:06 PDT):

>yep signoff and that we have a good sense that the deal will go through

Chaya Nayak (3/13/2024 16:24:22 PDT):

>but maybe that's its own stage

Chaya Nayak (3/13/2024 16:24:31 PDT):

>maybe we should delete that and just note assessment completed

Shelley Venus (3/13/2024 16:24:51 PDT):

>i think we can give you targets and costs at the same time

Shelley Venus (3/13/2024 16:25:10 PDT):

>and change that middle column to internal assessment/alignment

Chaya Nayak (3/13/2024 16:25:18 PDT):

>cool that sounds good

Shelley Venus (3/13/2024 16:25:21 PDT):

>@Brian Gamido you good with that

Chaya Nayak (3/13/2024 16:25:25 PDT):

>thank you!!

Brian Gamido (3/13/2024 16:28:11 PDT):

>yes. added estimates for stock clips

Shelley Venus (3/13/2024 16:29:07 PDT):

>would it be safe to say that we could get internal sign offs within a few weeks for the non-controversial categories @Chaya Nayak or do we need to incorporate time for an extraordinary budget ask

Shelley Venus (3/13/2024 16:36:18 PDT):

>(I am just defaulting to 4 weeks from now for the time being)

Brian Gamido (3/13/2024 16:38:25 PDT):

>@Chaya, do you have more clarity on the audio/stock line item. We've acquired sound effects data in the past and also acquired emotive voice data for fuchun in the past as well. Is that the right scope for that line?

Shelley Venus (3/13/2024 16:40:43 PDT):

>I added it @Brian Gamido yes to capture those types of providers (vs consumer facing audio dialogue i.e. podcasts).

Brian Gamido (3/13/2024 16:42:19 PDT):

>got it. i can add estimates for thinks like sound effects and emotive voice

Chaya Nayak (3/13/2024 16:44:51 PDT):

>we can get sign off pretty quick

Chaya Nayak (3/13/2024 16:45:03 PDT):

>yes!! can we use any of that out of the box?

Chaya Nayak (3/13/2024 16:45:26 PDT):

>Shelley can we do the news review faster? / when can we do the news review :)

Shelley Venus (3/13/2024 16:47:01 PDT):

>ha sorry i was being overly cautious. we could do sooner I am just flagging that this one is pretty tricky. per the news review in Feb we noted that we actually keep Meta's BoD updated on changes to our news policies.

Shelley Venus (3/13/2024 16:47:55 PDT):

>So if we pivot our company position to begin paying for news, it just might take longer than the other categories to get the right folks aligned but we can certainly try!

Shelley Venus (3/13/2024 16:48:24 PDT):

>i just changed it. shoot for the moon

Shelley Venus (3/13/2024 16:49:22 PDT):

>a lawyer somewhere just died a little inside 😢

Didem Foss (3/13/2024 16:53:11 PDT):

>@Chaya, we are filling out the text category (thanks Shelley for already filling it out with likely candidates)

>

>@Chaya, want to flag for you: we looked at books and scientific articles last year, and the decision was not to move forward, and recently we looked quora, for example, in the social platform category, and the decision was again to pass.

>

>The approach here is that the current exercise may override past decisions, and we should revisit these categories, even if the decision was to pass in the past, right?

Chaya Nayak (3/13/2024 16:53:44 PDT):

>correct we might override past decisions

Chaya Nayak (3/13/2024 16:54:11 PDT):

shared: 363519422_219978361029841_5840576661070952203_n.gif

Chaya Nayak (3/13/2024 16:54:13 PDT):

>we need more data

Shelley Venus (3/13/2024 17:04:00 PDT):

>@Chaya Nayak - honestly most of the news deck from Feb is all still accurate. we were just too early to have a robust consensus on moving forward.

>

>We'd really just want to dust it off and add more product precision

Shelley Venus (3/13/2024 17:04:21 PDT):

><https://docs.google.com/presentation/>

Shelley Venus (3/13/2024 17:04:41 PDT):

>assume the same for some of the discussions you had last year didem

Chaya Nayak (3/13/2024 17:05:02 PDT):

>If we can get this convo moving quick for editorial content that would be really helpful

Chaya Nayak (3/13/2024 17:05:15 PDT):

>for model training in particular

Chaya Nayak (3/13/2024 17:05:20 PDT):

>not for serving real time news

Shelley Venus (3/13/2024 17:16:31 PDT):

>Also, fyi for you and Mike - just to close the loop in case it was not stated explicitly - legal is

Redacted

Redacted

Chaya Nayak (3/13/2024 17:18:28 PDT):
Redacted

Brian Gamido (3/13/2024 19:15:02 PDT):
Redacted

Chaya Nayak (3/13/2024 19:16:26 PDT):
>@Sean Bell we should have your team look at these

Chaya Nayak (3/13/2024 19:16:33 PDT):
>Brian can you make us a list?

Chaya Nayak (3/13/2024 19:25:03 PDT):
>"Defer to Sy's POV on timing for books, academic pubs etc. We froze the GTM mid-negotiation last year. A bit painful all around :)" **Redacted - Privilege**

Chaya Nayak (3/13/2024 19:25:10 PDT):
Redacted - Privilege

Shelley Venus (3/13/2024 19:26:13 PDT):
Redacted - Privilege

Shelley Venus (3/13/2024 19:26:47 PDT):
>Didem Foss Brian Gamido lived through this though! We hadn't made it far with any news publishers

Sean Bell (3/13/2024 19:27:19 PDT):
>sorry, long chat here - what link(s) am I looking at?

Chaya Nayak (3/13/2024 19:32:31 PDT):
>@Mike Clark **Redacted - Privilege**

Shelley Venus (3/13/2024 19:32:55 PDT):
>Sean Bell Brian is pulling a list of past ML deals

Chaya Nayak (3/13/2024 19:33:06 PDT):
>this!

Chaya Nayak (3/13/2024 19:33:21 PDT):
>tldr we have some audio already secured for another team that we could use

Sean Bell (3/13/2024 19:33:45 PDT):
>Sure let's collect it and bring it to the team. Note that audio is another team anyway so they might be aware of it.

Sean Bell (3/13/2024 19:33:52 PDT):
>My team is not *yet* doing audio

Sean Bell (3/13/2024 19:33:56 PDT):
>(we are about to)

Mike Clark (3/13/2024 19:34:15 PDT):
>legal **Redacted**

Sean Bell (3/13/2024 19:38:14 PDT):
Redacted

Brian Gamido (3/13/2024 19:38:53 PDT):
>as a preview for voice data, here's what was acquired for embodiment work (merlyn deng, fuchun peng) last quarter:

Brian Gamido (3/13/2024 19:38:54 PDT):

shared: 431597880_3770001743253513_7986028539968176578_n.png

Brian Gamido (3/13/2024 19:45:59 PDT):
>'current status'

Sean Bell (3/13/2024 20:37:28 PDT):
>Question - why do we think we can't get licenses this month for, say, RawFilm or companies like that?

Sean Bell (3/13/2024 20:37:47 PDT):

shared: 418596861_367281262805262_625167500481097423_n.png

Sean Bell (3/13/2024 20:38:03 PDT):

>Feels like unless there's a legal problem, we should be able to buy the data in days, if Ahmad says "go"?

Sean Bell (3/13/2024 20:39:38 PDT):

>We put "5/31 license secured" for stock footage

James Kohli (3/13/2024 20:39:58 PDT):

>Yeah that column is supposed to be "here's when we'll have the timeline for acquisition ready to review, not here's when we'll have data in hand"

James Kohli (3/13/2024 20:40:17 PDT):

>Range on acquiring is so wide that I'm not sure it's a useful number

Sean Bell (3/13/2024 20:41:31 PDT):

>Also for TV studios, we already escalated this to Ahmad and he replied:

<https://docs.google.com/document/> [REDACTED]

James Kohli (3/13/2024 20:41:32 PDT):

>I agree with you, I think we've completed 90% of those columns for movies (minus a few around the long tail)

James Kohli (3/13/2024 20:42:29 PDT):

>I gotta go to bed, but I'll pick this up tomorrow. Video is in good shape to start acquiring, it's audio that's the big unknown

Sean Bell (3/13/2024 20:43:32 PDT):

>that's even worse 😞

Chaya Nayak (3/13/2024 20:45:20 PDT):

>these came from Shelly - let's try to push!!

Sean Bell (3/13/2024 20:46:00 PDT):

>to be fair to BD, eng was not providing a very clear "buy exactly this thing, right now, and here is budget" direction - but we now have that at least for some providers

Brian Gamido (3/13/2024 21:21:50 PDT):

>I'd say the wild card is biometrics consents--we do need smaller providers (and maybe also big providers) to get new consents from all subjects in the content or geo-fence their data by state

Brian Gamido (3/13/2024 21:22:59 PDT):

>*we may need